Remarks

Claims 1-7 are pending in the application. Claims 1-7 are rejected. All rejections are traversed. Claim 3 is canceled.

The specification is objected to. The specification has been amended to correct the informalities indicated by the Examiner.

> 7. Claims 1 and 2 are rejected under 35 U.S.C. 103(a) as being unpatentable over Shibata et al., (US 20020061136 A1), hereinafter Shibata, and further in view of Choi, (US 20020042793

The claims, as amended, claim updating a hierarchical hidden Markov model for each set of features extracted from a video.

The Examiner states:

> updating a hierarchical statistical model for each set of features [para. 0031, Fig. 1 is a schematic view showing a hierarchical model of video data; and para. 0052, see Fig. 1];

With all due respect, Fig. 1 in Shibata only shows a hierarchical structure of a video:

> "[0052] In the present invention, it is assumed that video data of an object of processing has such a modeled data structure as shown in FIG. 1 wherein it has three hierarchical layers of frame, segment and scene. In particular, the video data is composed of a series of frames in the lowermost hierarchical layer. Further, the video data is composed of segments, each of which is formed from a series of successive frames, in a higher hierarchical layer. Furthermore, the video data is composed of scenes, each of which is formed from segments collected based on a significant relation, in the highest hierarchical layer."

Shibata does not show a hierarchical hidden Markov model (HHMM) as claimed. Frames, segments and scenes are not hierarchical hidden Markov models.

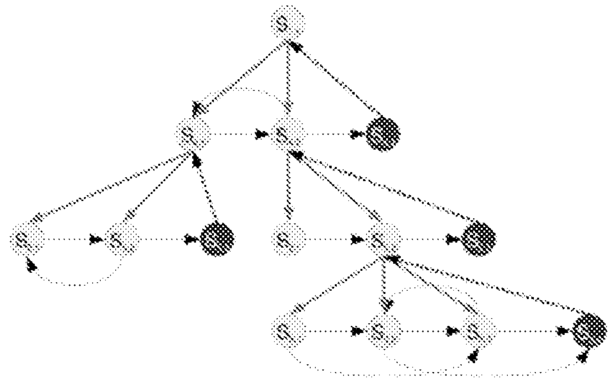
According to the MPEP, "During patent examination, the pending claims must be 'given their broadest reasonable interpretation *consistent with the specification.*' The Federal Circuit's en banc decision in *Phillips v. AWH Corp.*, 415 F.3d 1303, 75 USPQ2d 1321 (Fed. Cir. 2005) expressly recognized that the USPTO employs the 'broadest reasonable interpretation' standard… Indeed, the rules of the PTO require that application claims must 'conform to the invention as set forth in the remainder of the specification and the terms and phrases used in the claims must find clear support or antecedent basis in the description so that the meaning of the terms in the claims may be ascertainable by reference to the description.' 37 CFR 1.75(d)(1)… During examination, the claims **must** be interpreted as broadly as their terms reasonably allow. *In re American Academy of Science Tech Center*, 367 F.3d 1359, 1369, 70 USPQ2d 1827, 1834 (Fed. Cir. 2004) (The USPTO uses a different standard for construing claims than that used by district courts; during examination the USPTO ***must*** give claims their broadest reasonable interpretation **in light of the specification**.)."

In the above, emphasis is used to indicate that the Examiner **must** interpret the claims in light of the specification. The MPEP does not say that the Examiner is permitted to interpret the claims according to the cited art.

Formally defined, a hierarchical hidden Markov model (HHMM) is a statistical model derived from the hidden Markov model (HMM). In a HHMM, each state is considered to be a self-contained probabilistic model. More precisely, each state of the HHMM is itself a HHMM, which makes it hierarchical, see S. Fine, Y. Singer and N. Tishby, "The Hierarchical Hidden Markov Model: Analysis and Applications," Machine Learning, vol. 32, p. 41-62, 1998, and below:

We now give a formal description of an HHMM. Let Σ be a finite alphabet. We denote by Σ^* the set of all possible strings over Σ. An observation sequence is a finite string from Σ^* denoted by $\bar{O} = o_1 o_2 \cdots o_T$. A state of an HHMM is denoted by q_i^d $(d \in \{1, \ldots, D\})$ where i is the state index and d is the hierarchy index. The hierarchy index of the root is 1 and of the production states is D. The internal states need not have the same number of substates. We therefore denote the number of substates of an internal state q_i^d by $|q_i^d|$. Whenever it is clear from the context, we omit the state index and denote a state at level d by q^d. In addition to its model structure (topology), an HHMM is characterized by the state transition probability between the internal states and the output distribution vector of the production states. That is, for each internal state q_i^d $(d \in \{1, \ldots, D-1\})$, there is a state transition probability matrix denoted by $A^{q^d} = (a_{ij}^{q^d})$, where $a_{ij}^{q^d} = P(q_j^{d+1}|q_i^{d+1})$ is the probability of making a horizontal transition from the ith state to the jth, both of which are substates of q^d. Similarly, $\Pi^{q^d} = \{\pi^{q^d}(q_i^{d+1})\} = \{P(q_i^{d+1}|q^d)\}$ is the initial distribution vector over the substates of q^d, which is the probability that state q^d will initially activate the state q_i^{d+1}. If q_i^{d+1} is in turn an internal state, then $\pi^d(q_i^{d+1})$ may also be interpreted as the probability of making a vertical transition: entering substate q_i^{d+1} from its parent state q^d. Each production state q^D is solely parameterized by its output probability vector $B^{q^D} = \{b^{q^D}(k)\}$, where $b^{q^D}(k) = P(\sigma_k|q^D)$ is the probability that the production state q^D will output the symbol $\sigma_k \in \Sigma$. The entire set of parameters is denoted by

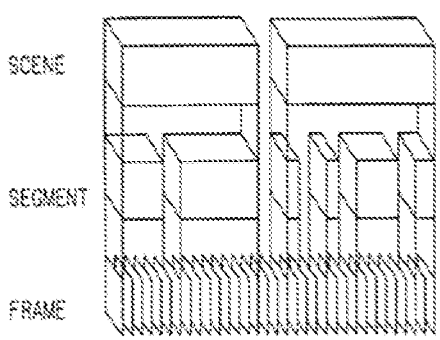
$$\lambda = \{\lambda^{q^d}\}_{d\in\{1,\ldots,D\}} = \left\{\{A^{q^d}\}_{d\in\{1,\ldots,D-1\}}, \{\Pi^{q^d}\}_{d\in\{1,\ldots,D-1\}}, \{B^{q^D}\}\right\} .$$

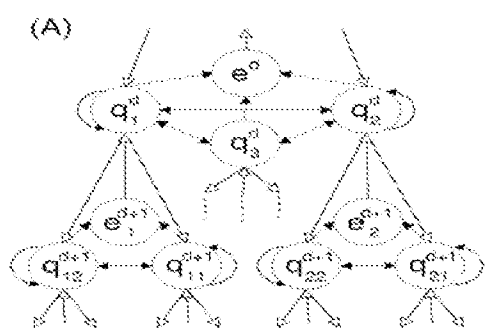


HHMM

Therefore, Shibata cannot evaluate an information gain of each feature on a Viterbi state sequence with respect to a reference state sequence of the hierarchical statistical model.

Applicants find the Examiner's argument that Shibata shows the claimed HHMM limitations unpersuasive. To assist understanding the difference, the cited art and the invention are distinguished below.





Shibata's hierarchy

Invention's HHMM

The Examiner states:

> evaluating an information gain of the hierarchical statistical model [para. 0104, corresponding to 'extracts a statistic representative value of an entire segment … n-dimensional vector … histogram and a power spectrum are involved; para. 0105 - 0106];

Applicants cannot find any reference to "a statistic… n-dimensional… histogram… power spectrum" in the reference paragraphs:

> "[0105] In the first case, the sample number is determined to be k in advance, and the video-audio processing apparatus uses a well-known k-means-clustering method disclosed in L. Kaufman and P. J. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis", John-Wiley and sons, 1990 to automatically divide the feature amounts regarding the entire segment into groups each including k feature amounts. Then, the video-audio processing apparatus selects, from each group of k samples, a sample whose sample value is equal or proximate to a centroid of the group. The complexity of the processing by the video-audio processing apparatus increases merely linearly in proportion to the sample number.
>
> [0106] Meanwhile, in the second case, the video-audio processing apparatus uses a k-medoids algorithm method disclosed in L. Kaufman and P. J. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis", John-Wiley and sons, 1990 to form groups of k samples. Then, the video-audio processing apparatus uses, as a sample value for each of the groups of k samples, a medoid of the group described above."

Furthermore, there is nothing in the claim about "a statistic…n-dimensional…histogram…power spectrum." The Examiner's is non sequitur.

Next, the invention claims filtering redundant features. The Examiner cites paragraph [0101]:

> "[0101] The video-audio processing apparatus extracts more than one static feature amount from different points of time within one segment, for example, as seen from FIG. 5. In this instance, the video-audio processing apparatus determines the extraction number of feature amounts by balancing maximization

> of the fidelity and minimization of the data redundancy in the segment representation. For example, where a certain one image in a segment can be designated as a key frame of the segment, a histogram calculated from the key frame is used as sample feature amounts to be extracted."

The meaning of filtering redundant features is plan and simple. But certainly neither "balancing maximization of the fidelity" nor "minimization of the data redundancy in the segment representation" have anything to with filtering redundant features in a HHMM. With all due respect, the Examiner's analogy makes no sense.

The invention claims updating the HHMM model based on the filtered features. The Examiner again cites paragraphs [0104-0107]. Forming a dissimilarity measurement for a feature does not update a HHMM as claimed. Furthermore, Shibata has no statistical model to update.

Shibata does not apply a Bayesian transformation to each HHMM.

There is no support in the Office Action for the statement:

> rank ordering the model and feature set pairs to learn the structure and detect the events in the video in an unsupervised manner" supra [as detailed].

Applicants cannot find anything 'supra' in the Office Action that details "rank ordering the (HHMM) model and feature set pairs to learn the structure and detect the events in the video in an unsupervised manner, rank ordering of a statistical model, and event detection in the Officer Action, or in Shibata.

The Examiner's comment, supra, is merely conclusionary. As is recognized in MPEP 707.07(d), "omnibus rejection of the claim ...is usually not informative and should therefore be avoided." MPEP 707.07(f) further mandates that "where a major technical rejection is proper, it should be stated with a full development of the reasons rather than by a mere conclusion coupled with some stereotyped expression." The rejection by the Examiner is a mere conclusion, without a full development of reasons. MPEP 706.07 further makes clear that "the invention as disclosed and claimed should be thoroughly searched in the first action and the references should be fully applied." In the present application, the rejection fails not only to provide a reasonable rational as to how, in the Examiner's view, the applied art can be construed to teach each and every feature in the rejected claims, but the rejection also fails to even consider explicitly claimed features of the invention as recited in claim 1.

From the above, it is clear that Shibata does not describe any of the claimed limitations. Therefore, Shibata cannot make the invention obvious Choi does not describe hierarchical hidden Markov models. Therefore, Shibata in combination with Choi can also not make the invention obvious.

Choi does not teach applying Bayesian information criteria to each HHMM and feature set pair. Choi describes using Bayesian techniques for self-organizing feature maps in the form of a neural network. There is no evidence that the Choi Bayesian techniques can be applied to HHMMs as claimed.

Choi deals with organizing documents. Choi is irrelevant for the problem faced by Shibata, namely determining a structure of a video. A person or ordinary skill in the art would never consider Choi as pertinent art to solve this problem. A combination of Shibata and Choi is technically impossible. Choi can never fail the numerous fatal defects of Shibata.

With respect to claim 2, the Gaussian distributions in Choi apply to network parameters. [0029] In a preferred embodiment of the present invention, the prior information determined in the sixth step takes the form of probability distribution, and the network parameter has a Gaussian distribution. Again, neither Shibata nor Chois have HHMMs that can be expressed as Gaussian distributions. A combination of Shibata and Choi is not feasible.

> 8. Claims 3 and 4 are rejected under 35 U.S.C. 103(a) as being unpatentable over Shibata et al., (US 20020061136 A1), hereinafter Shibata, in view of Choi, (US 20020042793 A1), and further in view of Lin et al., (US 7,076,102 B2), hereinafter Lin, or Ozer et al., (US 2004/0120581 A1), hereinafter Ozer, and/or Sterz, Walter et al., (DE 10122212 C1), hereinafter Sterz.

With respect to claims 3 and 4, neither Ozer nor Sterz update HHMMs from features extracted from a video. Lin deals with detecting events, not learning structure. Furthermore, Lin is inappropriate because in Lin the events are detected from objects that have been segmented from the video. The example that Lin uses is a falling body, and the object parts are the person's head, torso or legs. Obviously, Lin is not a feature-based system. Sterz deals with speech recognition, which is totally irrelevant to the problem of

Shibata. Furthermore, speech and video signals are completely different and techniques such as Sterz cannot be applied to signals processed by Shibata.

With respect to claim 5, the Examiner rejects that claim on Offical Notice, but provides no support for this assertion. MPEP 2144.03 states:

> "It would not be appropriate for the examiner to take official notice of facts without citing a prior art reference where the facts asserted to be well known are not capable of instant and unquestionable demonstration as being well-known. For example, assertions of technical facts in the areas of esoteric technology or specific knowledge of the prior art must always be supported by citation to some reference work recognized as standard in the pertinent art. ...; *In re Eynde*, 480 F.2d 1364, 1370, 178 USPQ 470, 474 (CCPA 1973) ('[W]e reject the notion that judicial or administrative notice may be taken of the state of the art. The facts constituting the state of the art are normally subject to the possibility of rational disagreement among reasonable men and are not amenable to the taking of such notice.')."

To generate a HHMM from dominant color ratios, motion intensities, least-square estimates of camera translation, audio volumes, spectral roll-offs, low-band energies, high-band energies, zero-crossing rates (ZCR) is not capable of such a demonstration. Generating HHMM, in general, is an extremely esoteric technology and difficult area to understand, for example, see Examiner's interpretation of HHMMs in the Office Action. Assertions that certain procedures in the field are well known must be supported by evidence as required by the MPEP.

Furthermore, MPEP 2144.33 goes on to state:

> "If such notice is taken, the basis for such reasoning must be set forth explicitly. The examiner must provide specific factual findings predicated on sound technical and scientific reasoning to support his or her conclusion of common knowledge. See *Soli*, 317 F.2d at 946, 37 USPQ at 801; *Chevenard*, 139 F.2d at 713, 60 USPQ at 241. The applicant should be presented with the explicit basis on which the examiner regards the matter as subject to official notice and be allowed to challenge the assertion in the next reply after the Office action in which the common knowledge statement was made."

The Examiner sets forth no explicit reasoning as to why he believes generating HHMM from dominant color ratios, motion intensity, a least-square estimate of camera translation, audio volume, spectral roll-off, low-band energy, high-band energy, zero-crossing rate (ZCR) is well-known, specifically in light of what the Examiner beliefs to be an HHMM. As such, the Examiner is depriving the Applicants of the ability to challenge his assertion, which is clearly in violation of the directives of the MPEP. The Examiner is respectfully requested to either provide evidentiary support and to clearly state his rationale in support of his assertions or to withdraw his rejections.

10. Claim 6 is rejected under 35 U.S.C. 103(a) as being unpatentable over Shibata et al., (US 20020061136 A1), hereinafter Shibata, in view of Choi, (US 20020042793 A1), and further in view of Bremer et al., (US 2005/0176057 A1), hereinafter Bremer.

Claimed are features filtered with a Markov blanket. Bremer deals with mood disorders and manic depression in suicidal patents. It is highly unlikely, that Shibata or Choi would look to Bremer's mentally ill patents to solve the problem of determining a structure of a video. A combination of Bremer and Shibata would indeed be bizarre.

11. Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over Shibata et al., (US 20020061136 A1), hereinafter Shibata, in view of Choi, (US 20020042793 A1), and further in view of Altschuler et al., (US 6,012,052), hereinafter Altschuler.

Claimed is evaluating an information gain in a HHMM using expectation maximization and a Markov chain Monte Carlo method.

Again, the Examiner reaches for incompatible art, which has not relationship to the primary references. First note, that neither Shibata not Choi have HHMMs, neither does Altschuler. In Altschuler, a Markov chain is used to sample a joint distribution of resource transition probabilities in a network, not to evaluate an HHMM of a video. The EM algorithm is used to estimate free parameters.

The Examiner states:

> Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to apply 'feature amounts, hierarchical model of video data, statistic representative value of an entire segment and minimization of the data redundancy disclosed by Shibata in combination with 'Bayesian self-organizing feature maps (SOM)', and 'Bayesian statistical technique' disclosed by Choi, and motivated to combine the teachings because Shibata is directed toward audio visual signal processing of video data, and although Choi is directed toward performing real-time document clustering for relevant documents in accordance with a degree of semantic similarity , Choi would improve the accuracy of information retrieval via performing real-time clustering for relevant information as revealed in para. 0002, and further coupled with 'Markov Chain Monte Carlo algorithms' and ''Expectation Maximization (or "EM") algorithm' disclosed by Altschuler and motivated to couple the teachings because the methods to pre-fetch resources and build resource link topology templates may also be used for collaborative filtering as disclosed by Altschuler in col. 3, lns. 60-67.

As best as can be understood, the Examiner states that it would be obvious to combine Shibata with Choi and Altschuler, because the Choi Bayesian techniques for rank ordering Web pages and Alschuler's EM algorithm for making consumer preferences using collaborative filtering would improve an understanding of the structure the Shibata video. This is indeed interesting but unpersuasive, and with all due respect makes no sense, clarification and amplification is respectfully requested.

It is believed that this application is now in condition for allowance. A notice to this effect is respectfully requested. Should further questions arise

concerning this application, the Examiner is invited to call Applicants' attorney at the number listed below. Please charge any shortage in fees due in connection with the filing of this paper to Deposit Account 50-0749.

Respectfully submitted,
Mitsubishi Electric Research Laboratories, Inc.

By

/Dirk Brinkman/

Dirk Brinkman
Attorney for the Assignee
Reg. No. 35,460

201 Broadway, 8th Floor
Cambridge, MA 02139
Telephone: (617) 621-7517
Customer No. 022199